Seven11 Assignment 1   **Due 11:59pm, Wednesday, September 22**

This assignment is required for students taking the 12-unit course only.

For the pairwise sequence alignment algorithms that we discussed in class, the score of an alignment with indels does not depend on where the indels are located within the alignment. In particular, the penalty for k adjacent indels is the same as the penalty for k indels scattered throughout the alignment. This is called a *linear* gap penalty.

The linear gap penalty does not reflect the fact that the gain or loss of an entire sequence fragment in a single event is less likely to disrupt protein structure than the gain or loss of individual residues spread across disparate locations. Alignments with a small number of large gaps are more readily obtained with an *affine gap penalty*, which includes a high cost for the first indel and a much lower cost for adding an indel to an existing gap. In this assignment, you will use an affine gap penalty function that has this property, described in detail in:

- J.C. Setubal and J. Meidanis, *Affine gap penalty functions*. Introduction to Computational Molecular Biology, pp 64-66.

  `https://www.cs.cmu.edu/~durand/03-711/Readings/eserdurand2.pdf`

This assignment asks you to align the same pair of sequences with an affine gap function and with a linear gap function, for comparison. An alignment template is provided for each question, consisting of a grid with cells for the alignment score and the traceback arrows. You can submit your solutions by filling in the template with a stylus and saving the annotated pdf. Alternatively, you can write your solution on a print-out of the template or a sheet of graph paper and upload an image of your handwritten solution. A third possibility is to use excel as a tool for recording a table; type your final answer into excel and save the result as a pdf. An excel template has been provided for that purpose.

Collaboration is allowed on this homework. You must hand in homework assignments individually. List the names of the people you worked with:

1. **Alignment with linear gap penalties:**

   (a) Compute the global alignment of "RISK" with "RANSACK", with the following scoring system. Matches: $M = 4$, mismatches: $m = -2$, gaps: $g = -5$. Show your alignment matrix with scores and traceback.

   (b) What is the score of the optimal alignment? How many different optimal alignments are there? Show them.

2. **Alignment with affine gap penalties:**

   (a) Recompute the global alignment of "RISK" with "RANSACK" with an affine gap function, scored as follows. Matches: $M = 4$, mismatches: $m = -2$, gap opening penalty: $h = -7$, gap extension penalty $g = -1$. Recall that cost of the first gap in a sequences of indels is $h + g$; the remaining gaps cost g. Show all three alignment matrices.

   (b) What is the score of the optimal alignment? How many different optimal alignments are there? Show them.